

Discussions

Stijn Vansteelandt

Department of Applied Mathematics and Computer Sciences

Ghent University, Krijgslaan 281 S9

9000 Ghent, Belgium

**email:* stijn.vansteelandt@ugent.be

1. Introduction

The problem of confounder selection is central to the analysis of essentially all observational studies, yet it has received remarkably little attention in the causal inference literature. I congratulate Wang, Parmigiani and Dominici - hereafter referred to as WPD - for addressing this important problem and making a very elegant proposal.

The problem of confounder selection differs from other covariate selection procedures in that confounders are, by definition, simultaneously associated with exposure and outcome. Yet, by factorization of the observed data likelihood into the outcome distribution (conditional on exposure and covariates) and the exposure distribution (conditional on covariates), standard procedures involve building models for these separate distributions one at a time. Methods solely based on outcome regression therefore have a tendency to ignore potentially important confounders. Namely, those covariates that are strongly associated with the exposure, but only moderately with the outcome, risk being dismissed from the analysis model as a result of multicollinearity. This may result in biased outcome regression analyses, which severely understate the actual uncertainty regarding the exposure's effect. In the causal inference literature, this problem has been mitigated through the development of propensity score methods which explicitly incorporate the exposure distribution into the analysis (Rosenbaum and Rubin, 1983). However, with few exceptions (e.g. McCandless, Gustafson and Austin, 2009), these procedures also separate model building for the propensity score from model building for the outcome. Contrary to outcome regression methods, they thereby prioritize covariates that are strongly associated with the exposure, regardless of their association with the outcome. It has been shown that this can result in inefficient inferences (Hahn, 2004). In addition, any bias in the exposure effect (e.g. due to unmeasured confounding, which is almost always present) gets amplified by the exposure's variance

inflation factor, which can be sizable when the model includes strong predictors of the exposure (Pearl, 2010; Vansteelandt, Bekaert and Claeskens, 2010).

All the above arguments call for simultaneously selecting covariates for inclusion in models for the outcome and exposure distributions. The authors' proposal prevents factorization of the joint outcome and exposure distributions through a prior dependence on the covariate inclusion probabilities. It forms a promising step in this direction.

2. How does BAC relate to propensity score adjustment?

For $\omega = \infty$, the proposed procedure imposes that covariates appearing in the exposure model will also be included in the outcome model, but allows for the outcome model to include more. This is also characteristic of the frequentist approach of Crainiceanu, Dominici and Parmigiani (2008). Not surprisingly, a similar performance is thus seen in the reported simulation studies. One important difference is that the use of model averaging may result in more honest descriptions of the overall model uncertainty. A further potential difference comes from the procedure's a priori tendency to exclude covariates from the exposure model, which can be seen because $P(\alpha_m^X = 0) = 2/3$ when $\omega = \infty$. From a subjective Bayesian viewpoint, $P(\alpha_m^X = 0)$ ought to represent one's prior beliefs. Considering that in most practical applications all covariates are - realistically seen - at least somewhat associated with the exposure and that the decision to measure data on a specific covariate in observational studies is often based on prior belief that this covariate is jointly associated with exposure and outcome, lower values of this probability could be desirable. In particular, it appears more prudent to choose a maximum value of $P(\alpha_m^X = 0) = 1/2$, which expresses a priori ignorance. For this choice, one would obtain

$$\frac{P(\alpha_m^X = 1 | \alpha_m^Y = 0)}{P(\alpha_m^X = 0 | \alpha_m^Y = 0)} = \frac{2}{\omega + 1}, \frac{P(\alpha_m^X = 1 | \alpha_m^Y = 1)}{P(\alpha_m^X = 0 | \alpha_m^Y = 1)} = \frac{2\omega}{\omega + 1}.$$

It would be interesting to see how this (or related) modification(s) might affect the performance of the proposed strategy, especially in settings where various covariates are strongly predictive of the exposure and only moderately predictive of the outcome.

The approach of Crainiceanu, Dominici and Parmigiani (2008), and thereby also BAC at $\omega = \infty$, has close connections with regression adjustment for the propensity score. This is so, even though the propensity score is usually defined for dichotomous exposures, because the propensity score can be redefined to be the conditional expectation of the exposure, given confounders, when the interest lies in linear regression models for the outcome (see e.g. Robins, Mark and Newey (1992)). A common strategy is then to build a propensity score model using standard model building steps, and to subsequently regress the outcome on the exposure, propensity score and those covariates that are most strongly associated with the outcome. Like BAC at $\omega = \infty$, this approach tends to adjust for all important predictors of the exposure (via the propensity score) in order to control for confounding, and additionally adjusts for strong predictors of the outcome in order to improve efficiency. It could nonetheless be preferable to BAC. First, because all predictors of the exposure are combined into a univariate propensity score, the outcome regression model will likely become more parsimonious. Second, unlike BAC, it entails a doubly robust procedure (Robins, Mark and Newey, 1992), which yields consistent additive exposure effect estimators when either the propensity score model or the outcome regression model is correctly specified. Alternatively, one may regress the outcome on the exposure and a model-averaged propensity score, as obtained via standard frequentist or Bayesian model averaging procedures. It will be worthwhile to formally compare BAC with these and other alternatives.

To gain preliminary insight, I have repeated simulation study 2 of WPD with $n = 100$ to mimic a setting where the information is scarce. The suggested propensity score procedure was implemented in two ways: once via stepwise regression (using the BIC) and once via a

model-averaged propensity score using the weights $\exp(\text{AIC}/2)$ (Buckland et al., 1997), where the AIC relates to the exposure model only. This resulted in an exposure effect estimator with bias 0.029 and 0.007, empirical standard deviation 0.216 and 0.218, and MSE 0.047 and 0.048, respectively. This improves upon fitting a main effects model with all covariates, but is inferior to BAC, which respectively yield a bias of 0.014 and 0.059, empirical standard deviation 0.255 and 0.162, and MSE 0.065 and 0.032. These differences appear at odds with the close relatedness of the different proposals. They call for a more in depth exploration to understand whether the superior performance of BAC is consistent (versus specific to this simulation) and whether it is related to its preferential tendency to exclude covariates from the exposure model, which could be a concern under other data-generating mechanisms with strong correlation between confounders and exposure.

3. Limitations of BAC

The proposals of WPD appear generic, but are essentially limited to linear models. This is because adjustment for a covariate in a non-linear model typically changes the meaning and magnitude of the exposure effect systematically. Contrary to what the authors claim, this is so even when that covariate is not associated with the exposure and thus not a confounder, as a result of so-called non-collapsibility of many association measures. Adjustment for such covariates in non-linear models is also typically disadvantageous as it may reduce precision (Robinson and Jewell, 1992). Vansteelandt, Bekaert and Claeskens (2010) overcome these concerns by focussing confounder selection on the quality of population-averaged exposure effects. For instance, in logistic regression models for the outcome,

$$E(Y_i|X_i, U_i) = \text{expit} \left(\beta^{\alpha_Y} X_i + \sum_{m=1}^M \alpha_m^Y \delta_m^{\alpha_Y} U_{im} \right),$$

they note that

$$E \{Y(x)\} = E \left\{ \text{expit} \left(\beta^{\alpha_Y} x + \sum_{m=1}^M \alpha_m^Y \delta_m^{\alpha_Y} U_{im} \right) \right\},$$

is the expected outcome that would be observed if all members of the population had exposure level $X = x$, provided that U_i is a set of covariates sufficient to control for confounding. The population-averaged exposure effect can then be defined as a contrast of $E\{Y(x)\}$ for different values of x , e.g. $E\{Y(x+1)\} - E\{Y(x)\}$ or $E\{Y(x+1)\} / E\{Y(x)\}$. Alternatively, it can refer to a contrast with the observed outcome, e.g. $E(Y) / E\{Y(0)\}$, or express the effect of changes in the observed exposure distribution, e.g. $E\{Y(1.1X)\} / E\{Y(X)\} = E\{Y(1.1X)\} / E(Y)$. Extending BAC/TBAC to these effect estimands will be essential.

A possibility would be to adapt the proposed procedure so that it provides model-averaged predictions of exposure and outcome, which may subsequently be used as input for more general (possibly frequentist) causal effect estimators. For instance, for discrete X , a doubly robust estimator of $E\{Y(x)\}$ is

$$\frac{1}{n} \sum_{i=1}^n \frac{I(X_i = x)}{P(X_i | U_i)} \{Y_i - E(Y_i | X_i = x, U_i)\} + E(Y_i | X_i = x, U_i).$$

The proposed procedure could be used to obtain model-averaged estimators of $P(X_i | U_i)$ and $E(Y_i | X_i = x, U_i)$ in a way that primarily targets the inclusion of all confounders. These arguments more generally suggest that the authors' focus on a single regression coefficient in the outcome regression model may be limiting.

4. What is BAC targeting?

BAC intuitively makes sense, but its heuristic nature leaves some vagueness as to whether it enjoys any specific optimality properties. Vansteelandt, Bekaert and Claeskens (2010) develop model selection strategies which target minimal mean squared error of the exposure effect estimator. This can be a desirable goal, given practitioners' disproportionate focus on point estimates. Low MSE may also indirectly be targeted by regressing the outcome on the exposure and a model-averaged propensity score, with model-averaging weights given by the reciprocal of the variance of the exposure effect estimator in a propensity score adjusted

analysis. For illustration, I evaluated this in the simulation study of the authors' Web Table 1 (where the presence of covariates that are solely associated with the exposure is particularly challenging for propensity score analyses). This resulted in an exposure effect estimator with bias 0.007, empirical standard deviation 0.040 and MSE 0.0016. This greatly improves upon both BAC and fitting a model which includes all main effects, which gave bias 0.001 and 0.007, empirical standard deviation 0.072 and 0.078, and MSE 0.006 and 0.006, respectively.

A more prudent strategy may be to seek an honest reflection of the overall uncertainty, and thus to target confidence validity. This is most easily obtained by avoiding model selection to the extent possible, by working under 'large models. Budtz-Jørgensen et al. (2005) find that this may even lead to tighter confidence intervals than those obtained after model selection when model uncertainty is taken into account. For instance, fitting a full model (i.e. including all covariates) considered in simulations 1 and 2 of WPD, resulted in an exposure effect estimator with bias 0.0005 and 0.002, empirical standard deviation 0.046 and 0.051, and MSE 0.002 and 0.003, respectively; coverages of 95% confidence intervals were 94.9% and 94.7%. These results are almost identical to those obtained by fitting the true model (see Tables 3 and 4, respectively), suggesting that there was actually no need for model selection in those simulation studies. To fully appreciate the performance of BAC, further simulation studies will be needed in settings where model reduction is essential.

Vansteelandt, Bekaert and Claeskens (2010) suggest alternative ways to guarantee confidence validity. Under certain assumptions, they demonstrate that a conservative asymptotic variance of the exposure effect estimator can be obtained, even when imprecision due to estimation and model selection on the propensity score is ignored, provided that the propensity score is efficiently estimated. Application of (such) procedures which exclusively adopt model building on the propensity score, thus guarantees (approximate) confidence validity, provided of course that the candidate list of models includes the true propensity score model.

It appears that BAC might come closer to a procedure that guarantees confidence validity for large ω (by forcing predictors of the exposure in the outcome regression model), although its potential tendency to preferentially exclude covariates from the exposure model leaves doubts as to whether this is generally true. For small ω , BAC might instead come closer to a procedure that favors minimal MSE, although it remains unclear to what extent this is so as BAC does not directly target minimal MSE. Further investigation will be important to understand the properties of BAC in function of ω . This will help practitioners to make an informed choice of ω , and may reduce the risk of selecting ω in a data-driven way.

5. Conclusions and possible extensions

In summary, the proposed procedure of confounder selection is attractive for its focus on confounders and for acknowledging model uncertainty. In its current form, the procedure is essentially limited to the analysis of linear models due to non-collapsibility of many association measures, but the approach could be extended to more general models along the lines described in Section 3.

The general principle of preventing factorization of the joint outcome and exposure distributions is attractive for confounder selection. It has also been seen in other contexts. For instance, when the outcome regression model is defined to be one that minimally includes the exposure and propensity score, then the inclusion of the propensity score necessarily prevents factorization (McCandless, Gustafson and Austin, 2009). Tan (2006) prevents factorization by ignoring part of the information on the joint distribution of outcome and covariates (at fixed exposure levels). It will be of interest to understand better how these different alternatives relate to BAC. For instance, what if standard Bayesian model averaging were applied on the basis of the joint outcome and exposure distribution in settings where the linear outcome regression model includes the propensity score?

The general principle of preventing factorization of the joint outcome and exposure dis-

tributions may also be more broadly applicable in other contexts. For instance, Greenland (2008) rightly argues that in practical applications, all available covariates will be somewhat associated with outcome and exposure. He therefore argues in favor of shrinkage estimators under a ‘large’ model for the outcome. A limitation of standard shrinkage procedures is that they do not target the exposure effect of interest. This may potentially be remedied by focussing on exposure and outcome models of the form

$$\begin{aligned}
 E(X_i|U_i) &= \sum_{m=1}^M \delta_m^X U_{im} \\
 E(Y_i|X_i, U_i) &= \beta X + \sum_{m=1}^M \delta_m^Y U_{im} \\
 \delta_m^X &\sim N(0, \sigma_X^2) \quad \text{for } m = 1, \dots, M \\
 \delta_m^Y | \delta_m^X &\sim N(0, \sigma_Y^2 (1 + \delta_m^{X2} / \sigma_X^2)) \quad \text{for } m = 1, \dots, M,
 \end{aligned}$$

where priors are used to penalize large regression coefficients. Here, the choice of priors ensures less penalization of outcome regression coefficients that correspond to strong predictors of the exposure. The resulting ridge regression appears close in spirit to BAC with a focus on confounding adjustment, but might be computationally advantageous relative to BAC and have the further advantage of not selecting confounders out of the analysis.

In conclusion, the principle underlying BAC holds much promise. However, as shown, many related - perhaps less computationally challenging - procedures can be conceived, which do not require novel model averaging strategies. Much obscurity remains as to how BAC relates to these and other more common approaches based on propensity scores and as to how to best choose ω in practice. Before widespread use of BAC can be recommended, closer scrutiny of its theoretical properties will be essential, as well as evaluation in realistic simulation studies where dimension reduction is critical (that is where, unlike in the considered simulation studies, fitting a full model which includes all covariates yields poorly performing estimators).

References

- Budtz-Jorgensen, E., Keiding, N., Grandjean, P. and Weihe, P. (2007). Confounder selection in environmental epidemiology: assessment of health effects of prenatal mercury exposure. *Annals of Epidemiology*, **17**, 27–35.
- Crainiceanu, C.M., Dominici, F. and Parmigiani, G. (2008). Adjustment uncertainty in effect estimation. *Biometrika* **95**, 635–651.
- Greenland, S. (2008). Variable selection versus shrinkage in the control of multiple confounders. *American Journal of Epidemiology*, **167**, 523–529.
- Hahn, J. (2004). Functional restriction and efficiency in causal inference. *Rev Econ Stat* **86**, 73–76.
- McCandless, L.C., Gustafson, P., and Austin, P.C. (2009). Bayesian propensity score analysis for observational data. *Statistics in Medicine*, **28**, 94–112.
- Pearl, J. On a class of bias-amplifying covariates that endanger effect estimates. In: Grunwald P and Spirtes P (eds) Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence, AOAI, Corvallis, OR, 2010, pp.417–424.
- Robins, J.M., Mark, S.D. and Newey, W.K. (1992). Estimating exposure effects by modelling the expectation of exposure conditional on confounders. *Biometrics*, **48**, 479–495.
- Robinson, L.D. and Jewell, N.P. (1991). Some Surprising Results about Covariate Adjustment in Logistic Regression Models. *International Statistical Review*, **59**, 227–240.
- Rosenbaum, P.R. and Rubin, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70**, 41–55.
- Tan, Z. (2006). A distributional approach for causal inference using propensity scores. *Journal of the American Statistical Association* **101**, 1619–1637.
- Vansteelandt, S., Bekaert, M. and Claeskens, G. (2010). On model selection and model misspecification in causal inference. *Statistical Methods in Medical Research*, November

12, 2010, doi:10.1177/0962280210387717.